110. Next Generation Multiplex Genome Engineering

Emily F. Freed* (emily.freed@colorado.edu) and **Ryan T. Gill**

University of Colorado, Boulder

Project Goals: This project is directed at the development of the next-generation of genome-engineering approaches to bridge predictive design with powerful methods for genome-scale search and optimization. Our studies combine iterations of detailed rational design, directed searches of mutational space, and high throughput mapping of protein sequence to activity. Our approach involves two objectives: i) to demonstrate an integrated, genome-scale strategy for "chassis" strain design and construction, and ii) to demonstrate a framework for "genome re-design" built upon multiplex synthetic biology and genome-engineering technologies. We have chosen to develop this platform by engineering i) ethylene production and ii) isobutanol production in *Escherichia coli*. Ethylene and isobutanol are attractive models because both can be directly converted to gasoline (and other advanced liquid biofuels) using well established catalytic routes and both have been produced in *E. coli* at various levels.

We are developing a set of tools that provides improved capabilities for linking genetic alterations to traits. Previously, the Gill laboratory developed trackable multiplex recombineering (TRMR), a technique that allows all genes in the *E. coli* genome to be barcoded and then simultaneously overexpressed or knocked down [1]. Cell populations containing barcoded expression level variants are then subjected to selective pressure and gene variants that result in increased fitness are identified by hybridization of their barcodes to a microarray. To date, TRMR has been used to map genes required for growth in various types of media and to optimize tolerance to acetate, low pH, cellulosic hydrolysate, isobutanol, ethanol, isopentenol, furfural, and various antibiotics [unpublished results, 1, 2]. These studies have given insight into carbon source and vitamin utilization, primary and secondary metabolism, and mechanisms of toxicity under a variety of conditions.

This study focuses on improving the TRMR design. The next generation TRMR library will be similar to the original library in that synthetic DNA (synDNA) cassettes will be used to replace the native promoter and RBS of a gene with a synthetic version of each, thus controlling gene and protein expression. However, the new library will allow for all *E. coli* genes to be expressed at four different levels (off, weak, intermediate, and strong), and will use a bicistronic RBS design (BCD) for more consistent expression levels across different genes [3]. Furthermore, an inducible promoter will be placed in front of each BCD allowing for fine tuning of gene expression to almost any level that is desired just by changing the amount of inducer that is added to the growth medium.

A second change will be the way in which relative fitness data is collected for each allele. In the original TRMR library, relative fitness was determined by hybridizing molecular barcodes to a microarray. In the next version of TRMR, molecular barcodes that are optimized for high throughput sequencing (e.g. Illumina HiSeq or Illumina MiSeq) will be used instead to track alleles. High throughput sequencing allows more quantitative analysis of genotype frequencies, since individual alleles will be tracked at the nucleotide level rather than by relative hybridization intensity (measured in arbitrary fluorescence units). A single run of Illumina HiSeq can generate 10^8-10^9 sequencing reads (a typical microarray signal distribution ranges over about 10^3), allowing for each barcode to be sequenced thousands of times. Another advantage is that barcodes for high throughput sequencing can be designed to be shorter in sequence length than those used for microarray, which reduces the cost of synthesizing the oligo library.

Finally, the new TRMR library is designed to integrate more easily with multiplex automated genome engineering (MAGE) [4]. One of the differences between original TRMR and MAGE is that original TRMR modifies both the promoter and the RBS while MAGE has only been used to modify the RBS. Next generation TRMR uses a synDNA cassette that can easily be modified for direct use in MAGE, which should lead to more consistent results between these two techniques and allow for even more rapid development of *E. coli* strains of interest.

The new TRMR libraries will be used in conjunction with the optimized chassis strains we are also developing to rapidly engineer and screen cell populations that have increased isobutanol or ethylene production.

References

1. Warner, J.R., et al., Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol, 2010. **28**(8): p. 856-62.
2. Sandoval, N.R., et al., Strategy for directing combinatorial genome engineering in Escherichia coli. Proc Natl Acad Sci U S A, 2012. **109**(26): p. 10540-5.
3. Mutalik, V.K., et al., Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods, 2013. **10**(4): p. 354-60.
4. Wang, H.H., et al., Programming cells by multiplex genome engineering and accelerated evolution. Nature, 2009. **460**(7257): p. 894-8.

This work is supported by the Office of Biological and Environmental Research in the DOE Office of Science.